Some Helpful Hints in Preparing Scientific-Quality Plots for Reports by hand or by using Excel

Gareth H. McKinley
2.671 Spring 2004

Excel is a wonderful tool for rapidly manipulating laboratory data 'on the fly', and extracting linear regression coefficients etc. However it is helpful to remember that it was designed with accountants in mind, not scientists! It is really configured to manipulate *financial* data, and this is reflected in the names used for many items. Data are referred to as 'series', graphs are referred to as 'charts', linear regression is referred to as a 'trend line' and extrapolating the results of a fit to higher/lower data values is referred to as 'forecasting' !! It takes some significant effort to make data tables and plots into the form that is acceptable for reporting scientific data. Even with this extra effort, the results are marginal at best (for example in my research group, students use Excel extensively for manipulating experimental data and displaying results in their lab books, but I forbid them from using plots made in Excel anywhere in their thesis or paper submitted for publication).

We wont take this restrictive step in 2.671, but unfortunately since Excel is frequently the first and only plotting tool students learn these days, the bad habits become inculcated. The following notes may help you address some of the most egregious shortcomings. Many of you were marked down in the Lab #2 report (in which you had to tabulate data and prepare several plots) based on these items. To illustrate this I use the example you were asked to prepare for the low pass filter. The resulting plot and table are shown overleaf.

Data Representation

• Never, *ever* join experimental data points by a line passing through the points. Remember your data are observations made at discrete points. You have absolutely no confidence in values between these points; there could be a spike, an inflection point, a discontinuity etc. You only have information where you measured it. If you then perform a linear (or nonlinear) regression, or overlay the results of a model prediction you can use a continuous line. Model lines should be extended/extrapolated across the full range of the plot (unless there is a good scientific reason why the model is invalid in some region).

Probably the worst offense in Excel, is that the default plotting style is not only that points are connected together, but this is done with a *spline*, i.e. a nice smooth line that goes through all the points and has a continuous derivative. This leads to nice smooth looking curves (always good for a financial analyst) but it misrepresents your data, and gives undue emphasis to statistical outliers. In Excel you should use the x-y scatter' plot and select the 'points only' icon.

• Always add a Legend' labeling all of the individual 'data series' plotted on the chart

Excel can help you in this regard. Many people didn't label the different data sets they plotted on the graphs in their report. You need to learn the options in Excel under the "Chart" menu for "Adding Data Series" etc.

• Tables: always use a preceding zero in the data you record in tables. Never switch units in the middle of tables (e.g. from mV to μV etc). Always use a consistent number of significant digits. This may be different column from column (e.g. if different instruments you have been recording from have different number of digit readout), but should be the same for all rows in a given table. It can be hard to do this in Excel by default (for example; if you use the 'general format' under 'Format: Cells' you get a fixed number of decimal places; e.g. 3.91, 0.39, 0.04….; this is NOT a consistent number of significant digits of course; better is to use the 'scientific notation' with a fixed number of significant digits (so the data series above would become 3.91E00, 3.91E-01, 3.91E-2….etc)

• always display UNITS in each column within Excel. This will help you when you go back and look at the data later. Mathcad is much better in this regard, but since they don't have a Mac version I can't use it. In Excel my convention is to always add another row or column that reminds me what the units for each column should be. Get into the habit of this and it will serve you in good stead later.

"Power User" Level

The above information constitutes a minimum level of effort that is required for good quality data and plots. If you are thinking of graduate school in the future (or even of excelling in industry!) then the following additional stylistic items will help you become an Excel power-user and bring it as close as possible to 'scientific quality'. You might like to reflect as you repeatedly click on charts in the future to set these options why almost all of these settings are the exact opposite of the defaults chosen by Excel!

- get rid of the default gray-shaded background (nice for business…never for science!)
- place tick marks on the inside of the axes, not the outside.
- In axes labels (and elsewhere) get into the habit of using [ ] for units *not* ( ). This allows you to write functions such as "$v_{meeasure}(t)$ [V]" more clearly! Also, variables should be *italicized.*
- use color to differentiate lines, but be smart and over-ride the default Excel colors; pick colors that display well on screen and paper (not yellow, light green etc). Start with primary then secondary colors (red, green, blue, then cyan, magenta….)
- If you are plotting multiple data series, alternate between hollow and filled symbols…

CONSTANTS

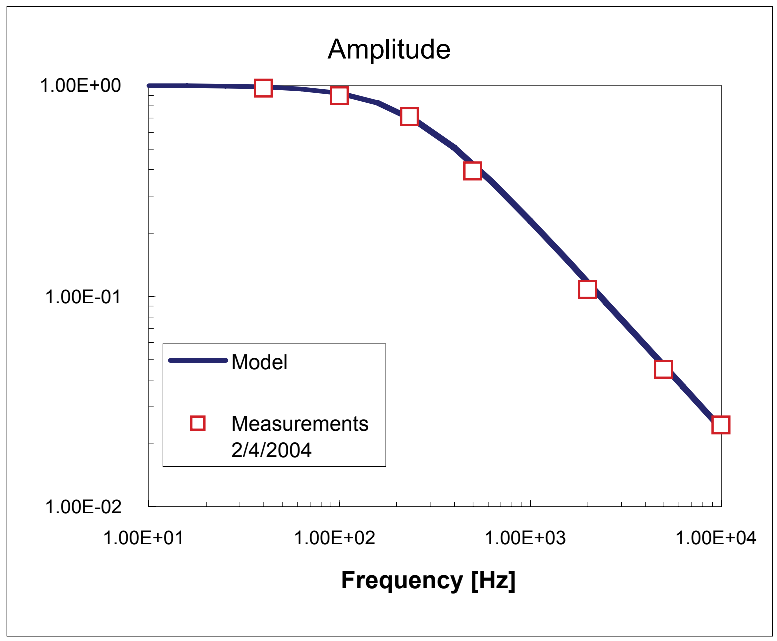
R	6.79E+00	kΩ
C	1.00E-01	μF
Characteristic Frequency		
f_c	2.34E+02	Hz

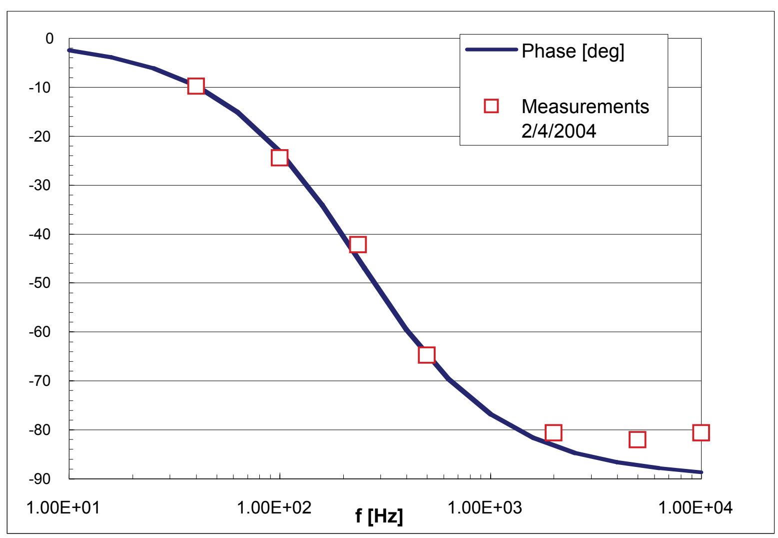
DATA

f [Hz]	Vout [V] *MEASURED*	Vout/Vin	t_f [ms] *MEASURED*	T [ms]	Phi [deg]
40	1.00E+00	9.70E-01	-6.80E-01	25.00	-9.79
100	9.22E-01	8.94E-01	-6.80E-01	10.00	-24.48
234	7.34E-01	7.12E-01	-5.00E-01	4.27	-42.19
500	4.06E-01	3.94E-01	-3.60E-01	2.00	-64.80
2000	1.11E-01	1.08E-01	-1.12E-01	0.50	-80.64
5000	4.60E-02	4.46E-02	-4.56E-02	0.20	-82.08
10000	2.52E-02	2.44E-02	-2.24E-02	0.10	-80.64

LOW PASS FILTER THEORY

f [Hz]	x	Model	Phase [rad]	Phase [deg]
1.00E+01	4.27E-02	9.99E-01	-0.042637	-2.44
1.58E+01	6.76E-02	9.98E-01	-0.0675133	-3.87
2.51E+01	1.07E-01	9.94E-01	-0.1067568	-6.12
3.98E+01	1.70E-01	9.86E-01	-0.1682383	-9.64
6.31E+01	2.69E-01	9.66E-01	-0.2629514	-15.07
1.00E+02	4.27E-01	9.20E-01	-0.403249	-23.10
1.58E+02	6.76E-01	8.28E-01	-0.5945463	-34.06
2.51E+02	1.07E+00	6.82E-01	-0.8199665	-46.98
3.98E+02	1.70E+00	5.07E-01	-1.0386704	-59.51
6.31E+02	2.69E+00	3.48E-01	-1.215104	-69.62
1.00E+03	4.27E+00	2.28E-01	-1.3405568	-76.81
1.58E+03	6.76E+00	1.46E-01	-1.4239667	-81.59
2.51E+03	1.07E+01	9.29E-02	-1.477751	-84.67
3.98E+03	1.70E+01	5.88E-02	-1.5119866	-86.63
6.31E+03	2.69E+01	3.71E-02	-1.5336641	-87.87
1.00E+04	4.27E+01	2.34E-02	-1.547361	-88.66





Notes:

(i) Notice the consistent number of significant digits and use of units…

(ii) To get n evenly spaced log points per decade, increase the x-value by $10^{1/n}$. e.g. I wanted 5 points per decade so the first column increases by $10^{0.2} = 1.585$.

(iii) Display of some fonts in the plots opposite may be imperfect as I have cut and pasted from Excel into Word and then 'shrunk to fit'.